The Angle between a random vector and its binarization.

Setup

$$b = \text{sign}(x) \in \{-1, +1\}^n$$

We want the angle θ between x and b, defined via:

$$\cos\theta = \frac{x \cdot b}{\|x\| \, \|b\|}$$

Key simplification

- $x \cdot b = \sum_i x_i \, \text{sign}(x_i) = \sum_i |x_i|$
- $\|b\| = \sqrt{n}$

So:

$$\cos\theta = \frac{\sum_i |x_i|}{\|x\| \sqrt{n}}$$

High-dimensional limit (main result)

If $x_i \sim \mathcal{N}(0,1)$, then by the law of large numbers:

- $\frac{1}{n}\sum_i |x_i| \to \mathbb{E}[|X|] = \sqrt{\frac{2}{\pi}}$
- $\frac{1}{n}\|x\|^2 \to 1 \Rightarrow \|x\| \approx \sqrt{n}$

So:

$$\cos\theta \to \sqrt{\frac{2}{\pi}} \approx 0.798$$

Hence:

$$\theta \approx \cos^{-1}(0.798) \approx 37^\circ$$

Interpretation

- The binarized vector b is **not orthogonal** to x
- It is **not far from perfectly aligned**
- The angle concentrates tightly around ~37° in high dimensions

This is a **concentration of measure** phenomenon: the angle becomes almost deterministic as $n \to \infty$.

Geometric intuition

- b keeps only the **orthant information** (which quadrant each coordinate lies in)
- x contains both **sign and magnitude**
- The loss of magnitude reduces alignment, but not drastically

You can think of this as:

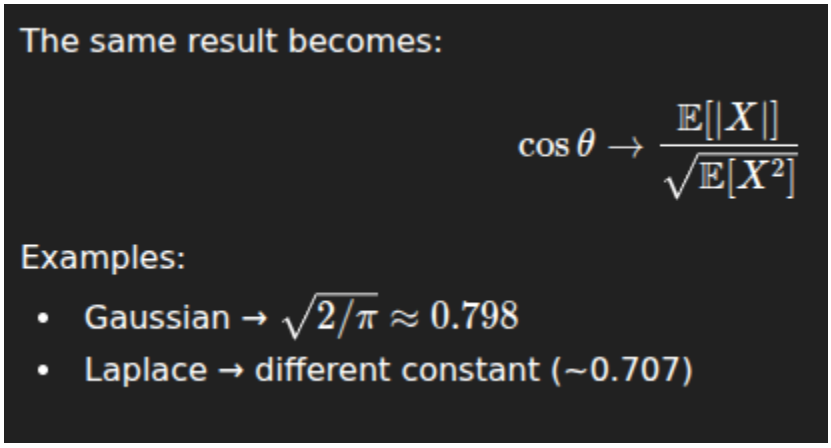
- projecting x onto the corners of the hypercube $\{-1,+1\}^n$
- the nearest corner still has a strong directional correlation

---

Generalizations

1. Other symmetric distributions

The same result becomes:


The same result becomes:
$\cos\theta \to \frac{\mathbb{E}[|X|]}{\sqrt{\mathbb{E}[X^2]}}$
Examples:
• Gaussian → $\sqrt{2/\pi} \approx 0.798$
• Laplace → different constant (~0.707)

Examples:

- Gaussian → 0.798
- Laplace → different constant 0.707

---

2. Relation to random hyperplanes

This connects to results in:

- random projection
- sign-based hashing (e.g., SimHash)

The sign operation preserves **angular information surprisingly well**, which is why binary embeddings work.

---

3. Neural network view

In a ReLU or sign network:

- this angle quantifies **information loss due to gating**
- even extreme binarization retains ~80% cosine similarity

This helps explain why:

- binarized networks can still work reasonably well
- but need depth or redundancy to recover lost magnitude information

---

Takeaway

- The angle between a random vector and its sign vector **converges to a constant**
- For Gaussian inputs:
  Angle ~37°
- So binarization preserves a **surprisingly strong directional signal**, despite discarding all magnitude information

---